Indian Journal of Technology
Vol. 16, April 1978, pp. 148-153

Vowel Identification Using Piecewise Separation Technique

S.K. PAL & D. DUTTA MAJUMDER
Electronics and Communication Science Department, Indian Statistical Institute, Calcutta 700 035

Received 5 *February* 1977; *accepted* 20 *August* 1977

A simple method for computer recognition of Telugu speech sounds irrespective of speakers is described. A vocabulary consisting of 871 Telugu words containing the ten vowels (/ə/,/a:/,/i/,/i:/,/u/,/u:/,/e/,/e:/,/o/ and /o:/) in constant-vowel nucleus-consonant (CNC) combination and uttered by three informants was selected as the testing material. Formant frequencies, F_1, F_2 and F_3 of the vowel sounds were extracted by spectrographic analysis. "Piecewise separation" using $(F_2—F_1)$ or (F_3/F_2) as parimary recognition parameters and (F_3/F_1) as a final recognition criterion were used for classification. From frequency distributions of these parameters for both shorter and longer vowel categories suitable boundaries have been selected. In the classification model, shorter and longer subgroup of a vowel have been pooled together in the same class and the overall recognition score is about 84%.

MANY attempts have been made to build a speech recognition system[1-5] such that vowel speech sounds could be recognized in almost all cases either from the conditional sampled data of speech waveform[2,6] or from the vowel formant frequencies with a suitable discriminant function[7,8]. In the second case, the acoustic parameters of vowels such as voice fundamental frequency (F_0) and formant frequencies (F_1, F_2, F_3, etc.) depend greatly on sex and age. It has been found that F_0 and F_3 do not bear as much of the information about phoneme identity as do F_1 and F_2; they are also speaker dependent and possess close correlation with age and sex. Every measurement on a vowel in $(F_0$-$F_3)$ plane corresponds to almost identical region irrespective of the kind of vowel[9]. For this reason, vowel identification for a number of informants does show an improved result when either F_3 or F_0 or both F_3 and F_0 in addition to F_1 and F_2 are taken into consideration.

In the present paper we have made an attempt towards the recognition of Telugu vowels irrespective of speakers by considering only the third formant F_3 in addition to F_1 and F_2. The classification technique is based on piecewise separation by a decision boundary determined from the distribution functions of the characterizing parameters. These parameters were found from the interrelations among F_1, F_2 and F_3. For threshold boundaries, $(F_2—F_1)$ or (F_3/F_2) were taken as the recognizing criteria in primary separation and (F_3/F_1) is the last criterion for final recognition. Their distribution functions after preprocessing were also studied. Finally, a suitable piecewise linear classification scheme with the minimum number of errors is described. The results of computer identification of vowels are presented in tabular form.

Design of Feature Space

The main task before designing a pattern classifier is the extraction of significant characterizing features which determine the invariant and common properties of a set of members associated with a pattern class in the sample space. To learn the distribution of events in the sample space for Telugu vowels, three selected adult male informants were allowed to record a number of discrete PB (phonetically balanced) speech units in CNC (consonant-vowel nucleus-consonant) form. CNC combination is taken because the consonants in a form connected to a vowel are responsible for influencing the role and quality of vowels. The IPA (International Phonetic Association) list for ten Telugu vowels is given in Appendix 1. We have not used any diphthong nucleus in the experiment. Spectrographic analysis of these phoneme on Kay Sonagraph (Model No. 7029A) indicated permanent record of the formant frequencies. Some of the analysed records were not included with their respective third formant F_3. They were allowed to have injected average third formant frequency $(F_3)_{av}$, computed over all members of that class of vowels for the particular informant.

The samples with their respective features F_1, F_2 and F_3 constitute a 3-dimensional feature vector space Ω_F, each dimension representing an invariant property of the event. All the significant informations available about an event could be expressed as a 3-dimensional feature vector

$$\mathbf{F} = \begin{bmatrix} F_1 \\ F_2 \\ F_3 \end{bmatrix}$$

the co-ordinate of which have numerical values indicating the amount of each property. Each event with a set of all measured features will, therefore, correspond to a single point in the feature vector space Ω_F.

Selection of Recognition Parameters

Parameter selection for the present scheme of vowel identification is governed by some basic linear relations within the formants F_1, F_2 and F_3. The idea of connecting two formants arises from the following facts. The formant frequencies provide information about the position of the speaker's articulatory organ such that these frequencies can change only as a result of an articulatory change affecting the dimensions of the various parts of the

vocal tract cavity system. Again the statistical average distance between formants is physiologically correlated with the total length of the vocal tract[10]. Larger the cavity length, smaller is the average formant spacing. Hence, associating the two in suitable form like F_3/F_1, F_2/F_1, F_3/F_2, $F_3 - F_1$, $F_2 - F_1$, $F_3 - F_2$, etc. representing physiological correlation with the size of the different cavity resonators may improve vowel recognition.

After a study of the several probable parameters it was found that the functions $(F_2 - F_1)$ and (F_3/F_2) lead to a satisfactory primary classification scheme. From frequency distributions of the parameters $(F_2 - F_1)$ and (F_3/F_2) linked with each feature vector **F** as in Figs 1 and 2 respectively, it is noticeable that a suitable threshold level could be placed so that all vowel classes including shorter and longer ones stand partitioned into two sets, namely, Group I={Back vowels : /∂/, /a:/, (/u/, /u:/) and (/o/, /o:/)} and Group II={Front vowels: (/i/, /i:/) and (/e/,/e:/)}. For other parameters $F_3 - F_1$ and $F_3 - F_2$ say, it was found that the corresponding distributions vary approximately from 1300 to 2150 Hz and 600 to 1500 Hz for /∂/ and /a:/, 1900 to 2900 Hz and 200 to 900 Hz for /I/, 1300 to 2500 Hz and 700 to 2100 Hz for /U/, 1600 to 2600 Hz and 400 to 1400 Hz for /E/ and 1650 to 2500 Hz and 1050 to 2300 Hz for /O/. These parameters, therefore, did not show any discriminating properties and further the distributions only in case of vowel/E/ were found to have central tendency.

Figs 1b and 2b show the distribution functions for the samples when the vowels /∂/ and /a:/ are considered to be one and longer and shorter subgroups are treated in the same class. Shorter and longer types of a vowel show a constant lag between their distribution functions as shown in Figs 1a and 2a. The final classification within each group is developed by comparing the criterion (F_3/F_1) associated with each sample **F**. This is in agreement with the distribution characteristics with respect to the front and back vowels shown in Fig. 3.

Back vowels with suitably defined boundaries could be classified chronologically as /A/, /O/ and /U/ according to low, medium and high values of (F_3/F_1) and similarly front vowels possessing low and high (F_3/F_1) magnitudes with respect to a threshold are treated as /E/ and /I/ respectively. The final classification based on the parameter (F_3/F_1) does not lead

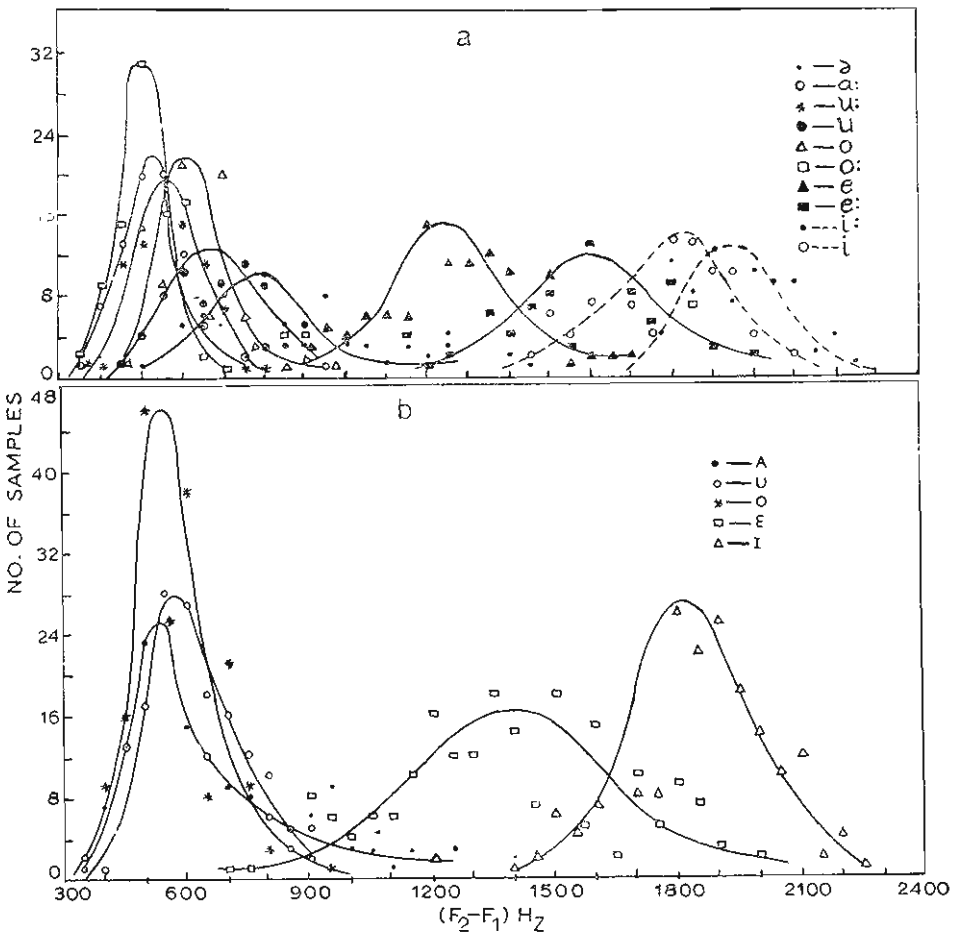

Fig. 1 — Frequency distribution of the parameter $(F_2 - F_1)$ when shorter and longer categories are (a) treated separately, and (b) pooled together

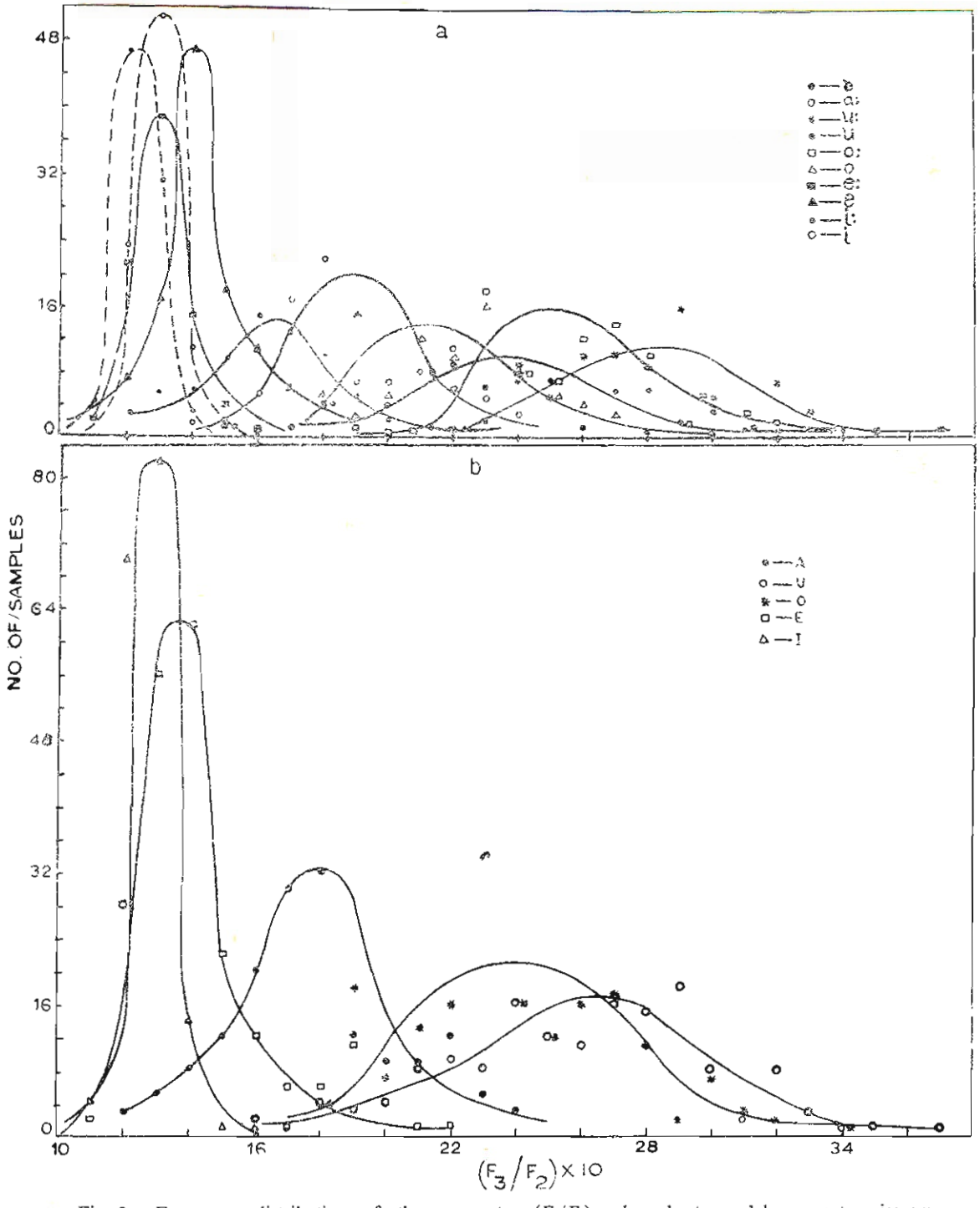

Fig. 2 — Freqnency distribution of the parameter (F_3/F_2) when shorter and longer categories are (a) treated separately, and (b) pooled together

to the isolation of vowel classes /∂/ and /a:/. These two vowel classes are better identified by applying the difference parameter $(F_2—F_1)$.

To have the knowledge about parameter distribution, they were suitably conditioned in two steps to provide convenient formant and also to provide invariance to the samples. At first, the magnitudes of the parameters (F_3/F_2) and (F_3/F_1) corresponding to the patterns were scaled up by a factor of 10 and finally they were subjected to an approximation to the nearest integers, whereas the difference parameter $(F_2—F_1)$ was processed to have the closest frequency value, which is a multiple of 50 Hz. The parameters thus processed make the classifier more effective.

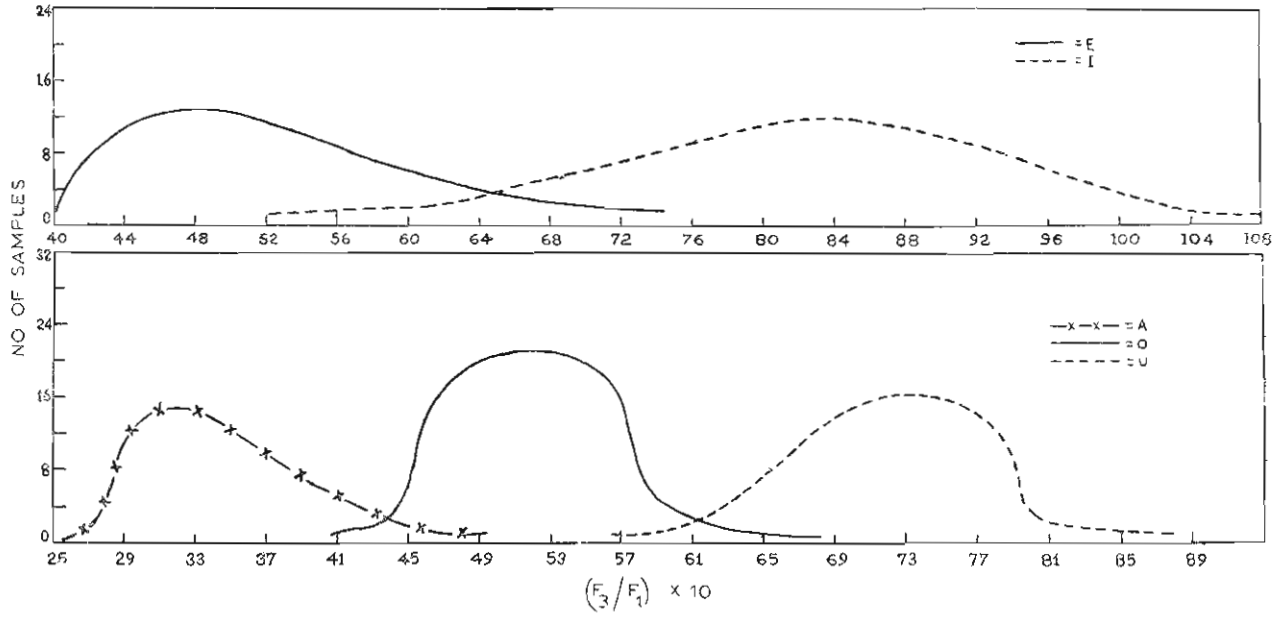

Fig. 3 — Frequency distribution of the parameter (F_3/F_1) when shorter and longer categories are pooled together

Classification Model

Consider an input pattern

$$F = \begin{bmatrix} F_1 \\ F_2 \\ F_3 \end{bmatrix}$$

represented by a point in the 3-dimensional feature vector space Ω_F, consisting of pattern classes C_{∂}, $C_{a:}$, C_i, $C_{i:}$, C_u, $C_{u:}$, C_e, $C_{e:}$, C_o and $C_{o:}$, associated with vowels /ə/, /a:/, /i/, /i:/, /u/, /u:/, /e/, /e:/, /o/, and /o:/ respectively to be recognized.

Before designing the classifier, the feature space Ω_F is subjected to the restriction that no two points defining the measured features of two different vowels having different time durations, but identical phonetical properties may lie within different classes and would be treated as member of the same, i.e.,

$$C_i, C_{i:} \subset C_I$$
$$C_u, C_{u:} \subset C_U$$
$$C_e, C_{e:} \subset C_E$$

and $$C_o, C_{o:} \subset C_O$$

where $\Omega_F = \{C_{\partial}, C_{a:}, C_I, C_U, C_E, C_O\}$

The function of the classifier to be designed is then to assign each input pattern F to its proper class by the piecewise separation technique, which is a two-fold task. In the first step, it would decide whether the pattern is a member of Group I or Group II denoted as Ω_{FI} and Ω_{FII} respectively, such that

$\Omega_F \supset \Omega_{FI}$ = {Back vowels : /ə/,/a:/,/U/and/O/} ...(1)

$\Omega_F \supset \Omega_{FII}$ = {Front vowels: /I/ and /E/} ...(2)

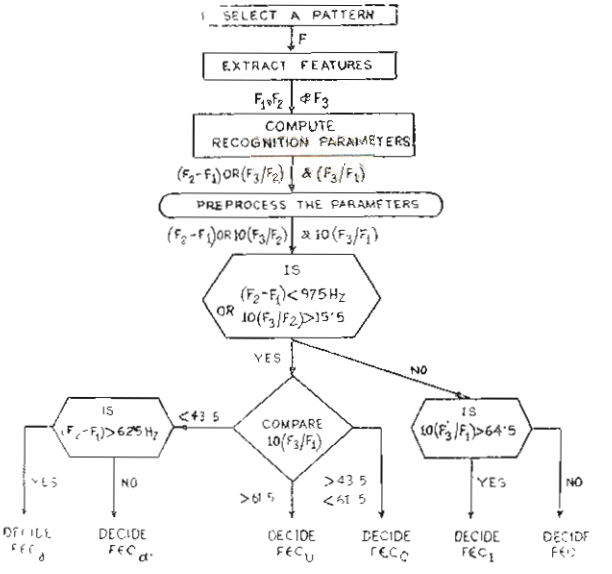

Fig. 4 — Block diagram representing classification of Telugu vowels

This could be reasonably solved from the frequency distribution curves corresponding to the parameters $(F_2—F_1)$ or (F_3/F_2) which argue that

$F \in \Omega_{FI}$, if $(F_2—F_1)<975$ Hz [or, $10(F_3/F_2)>15\cdot5$] ...(3a)

and $F \in \Omega_{FII}$, otherwise ...(3b)

The second step is to assign each pattern its proper class belonging to the primary selected group, on the basis of the parameter (F_3/F_1). The curves in Fig. 3 correspondingly indicate the following inequalities for final classification :

For Group I

If (i) $10(F_3/F_1)<43{\cdot}5$, decide $F\varepsilon$ C_∂ or $C_{a:}$ ...(4a)

in which C_∂ and $C_{a:}$ are recognized according to the condition, $(F_2 - F_1)$ is greater or less than 625 Hz.

(ii) $10(F_3/F_1)>61{\cdot}5$, decide $F\varepsilon$ C_U ...(4b)

and (iii) $61{\cdot}5>10(F_3/F_1)>43{\cdot}5$, decide $F\varepsilon$ C_O ...(4c)

For Group II

If (i) $10(F_3/F_1)>64{\cdot}5$, decide $F\varepsilon$ C_I ...(5a)

and (ii) otherwise, decide $F\varepsilon$ C_E .. ...(5b)

The piecewise recognition so programmed is shown in blocks in Fig. 4, where an input pattern $\mathbf{F}(F_1,F_2,F_3)$ with its extracted information and suitably processed parameters is finally subjected to go through the piecewise classification with a minimum probability of errors.

Experimental Procedure and Results

A vocabulary consisting of Telugu words for ten vowels was selected so as to encompass as many CN and NC combinations as possible with emphasis on the use of commonly used words. These were recorded by 5 adult male informants on an AKAI tape recorder inside a big auditorium. On the basis of a listening experiment by 10 listeners only 871 samples of three informants were selected. The spectrographic analyses of these utterances were done on a Kay sonagraph (Model 7029A). The analyses were carried out in the normal mode and the 0·08–8 kHz band with a wide band pass filter (bandwidth 300 Hz) had been chosen.

There are automatic algorithms[5,11] where insertion and deletion errors may be present due to segmentation of connected speech with acoustic boundaries into phonemic units. Since the present experiment is concerned with the recognition of only vowels from connected speech, the formants F_1, F_2 and F_3 were obtained manually at the steady state of the vowels.

The steady state of the vowel is that part on the record in which all formants lie parallel to the time axis. The transition is depicted by the inclined formant patterns. The exact point of inflection is difficult to locate in the records. This can be done satisfactorily by tracing the central line for each formant band. Once these points are located for all available formants, the steady state of the vowel is taken to be the shortest horizontal span for all the formants.

In view of the large amount of data to be handled, the formant frequencies have been measured from the base line with a specially constructed scale. A rechecking on 5% samples revealed that formant frequencies have been recorded accurate to within 10 Hz. In a few cases, for particularly fast informants, it has been noticed that the vowel hardly reaches a stable state. In such cases, the congruences of the on-glides and off-glides have been taken as the steady state. The samples which did not depict prominent third formant were allowed to have injected average third formant $(F_3)_{av}$, computed over all members of that class of vowels for that particular informant.

Table 1 — Average Formant Frequencies and Durations of Telugu Vowels

[The values are averaged over three male informants]

Vowel	F_1 Hz	F_2 Hz	F_3 Hz	$\triangle t$ msec
∂	606	1473	2420	98·53
a:	710	1240	2400	270·46
i	365	2116	2757	94·72
i:	325	2260	2836	257·80
u	370	1066	2500	98·23
u:	348	923	2543	268·03
e	517	1796	2633	112·96
e:	470	1883	2657	275·85
o	476	1133	2630	111·80
o:	486	1000	2540	244·00

Table 2 — Standard Deviations of First Three Formants of Telugu Vowels

Vowel	Standard deviation, Hz		
	First formant	Second formant	Third formant
∂	76·248	191·570	226·511
a:	60·560	97·006	155·760
i	75·367	185·209	237·048
i:	50·199	133·153	234·571
u	53·657	137·400	331·007
u:	35·560	68·032	124·411
e	81·775	201·020	197·481
e:	66·662	319·498	204·488
o	71·062	138·899	268·996
o:	44·357	66·862	231·343

The number of samples that fall in this category is 384.

The average formant frequencies and durations of Telugu vowels are shown in Table 1. The values given are averaged over three male informants. The duration includes the duration of the steady state as well as the duration of the on-glide and the off-glide. The data in Table 1 show that the shorter and longer vowels differ much in duration rather than in frequency range. The standard deviation for all the vowel formants is given in Table 2. Shorter categories are found to have larger variation than longer ones.

The parameters (F_2-F_1), (F_3/F_2) and (F_3/F_1) for each event were then computed and pre-processed, as described before to attain suitable rectified magnitudes required for plotting their distribution curves. Frequency distribution of parameters as plotted in Figs 1–3 are the mean curves drawn over the experimental points. Longer and shorter categories of vowels are observed to maintain a constant lag between their distribution curves (Figs 1a and 2a).

TABLE 3 — PERCENTAGE RECOGNITION OF TELUGU VOWELS

Vowel	No. of samples	Recognition score (%) when parameter (F_2-F_1) is taken	Recognition score (%) when parameter (F_3/F_2) is taken
∂	72	54·40	52·77
a:	89	85·40	86·51
i & i:	172	89·53	89·53
u & u:	151	90·06	90·06
e & e:	207	79·72	80·20
o & o:	180	89·44	89·44
Total :	871	83·69	83·35

The recognition percentage of each of the classes of Telugu vowels is shown in Table 3. It is seen that identification score for all the vowels except /∂/, is satisfactory. Poor recognition rate for C_o is because of the fact that members of this class as compared to other classes are much affected whenever a classification rule is adopted. With the same background events of the class C_E in Ω_{FII} is identified comparatively with lower score.

Either of the parameters (F_2-F_1) and (F_3/F_2) selected for primary separation of vowels in the two categories results in almost equal recognition. Though the error rates are almost same in both the cases, it is reasonable from the point of discrimination and time consumption, as seen from Figs 1 and 4 respectively, to consider the difference parameter (F_2-F_1) for primary classification.

Conclusion

The recognition rates for vowels /I/, /U/ and /O/ are found to be high (≈90%) compared to other classes and the total percentage recognition is about 83–84%. The algorithm presented here is, of course, restricted in treating the shorter and longer vowels, not differing phonetically but in duration only, in the same category. It also provides a much simpler and faster approach than other methods[7,9,12] for identifying vowels irrespective of speakers.

Acknowledgement

The authors thank Shri A.K. Datta for valuable suggestion and Sarvashri S. Chakraborty and S. Biswas for secretarial assistance. One of the authors (S.K. Pal) is also grateful to the Council of Scientific and Industrial Research, New Delhi for providing a fellowship.

References

1. DULLEY, H. & BALASHEK, S , *J. acoust. Soc. Am.*, **30** (1958), 721.
2. SAKAI, T. & DOSHITA, S., *IEEE Trans.*, EC **12** (1963), 835.
3. DUTTA MAJUMDER, D. & DATTA, A.K., *Proc. Automation and Instrumentation Conf , Milan, Italy*, (1968), 249.
4. PAL, S.K. & DUTTA MAJUMDER, D., *IEEE Trans, Systems Man and Cybernetics*, SMCT (1977), 625.
5. REDDY, D.R., *Proc. IEEE*, **64** (1976), 501.
6. LECOURSE, M. & SPARKERS, J.J., *Proc. Conf. Speech Commun. Process*, M.I.T., Cambridge, Mass., 1967, 64.
7. DUTTA MAJUMDER, D., PAL, S.K. & DATTA, A.K., *J. Computer. Soc. India*, **7** (1976), 14.
8. FU, K.S. *Sequential methods in pattern recognition* (Academic Press, New York), 1968.
9. SUZUKI, H., KASUYA, H. & KIDO, K., *Proc. Conf. Speech Commun. Process.*, M.I.T., Cambridge, Mass. USA, (1967), 92.
10. FANT, G., *Acoustic theory of speech production* (Mouton &Co., 'S—Gravenhage), 1960, 20.
11. REDDY, D.R., *J. acoust. Soc. Am.*, **40** (1966), 307.
12. DUTTA MAJUMDER, D. & PAL, S.K., *J. Inst. Electron. Telecom. Engrs*, **23** (1977), 117.

Appendix 1

IPA list for Telugu vowels

Telugu vowel	Phonetic symbol
అ	∂
ఆ	a:
ఇ	i
ఈ	i:
ఉ	u
ఊ	u:
ఎ	e
ఏ	e:
ఒ	o
ఓ	o: